

Application No. 10/044,711

Atty Docket No. INXT 1018-1

In the Specification:

Please amend paragraph [0047] as follows:

[0047] The nearest neighbors to the probe document have one or more topic assignments T_i . The category scores of the topics can then be calculated for the probe document in at least two ways:

$$\Omega_0(\mathbf{d}_t, T_m) = \sum_{\mathbf{d} \in \{K(\mathbf{d}_t) \cap T_m\}} s(\mathbf{d}_t, \mathbf{d}) \quad (0.1)$$

or:

$$\Omega_1(\mathbf{d}_t, T_m) = \frac{\sum_{\mathbf{d}_1 \in \{K(\mathbf{d}_t) \cap T_m\}} s(\mathbf{d}_t, \mathbf{d}_1)}{\sum_{\mathbf{d}_2 \in K(\mathbf{d}_t)} s(\mathbf{d}_t, \mathbf{d}_2)} \quad (0.2)$$

The first formula calculates the sum Ω_1 of the contributions or evidence that probe document \mathbf{d}_t should receive topic assignment T_m , while the second calculates the normalized sum Ω_2 , normalized to reflect the density of the neighborhood of the probe document. These formulas may be modified to reflect the number of documents among the nearest neighbors that share the topic assignment T_m . Figures 5A-B illustrate the significance of normalizing scores. In these figures, the triangles and circles represent documents having first and second categorizations, respectively. The density of the neighborhood around the test document, which is represented by the larger square, is indicated by the distance between the test document and its neighbors. An ambiguous situation, as in Figure 5A, may appear to strongly support a topic assignment, because the neighborhood 502 of the test document 501 is dense. An unambiguous situation, as in Figure 5B, may seem to weakly support a topic assignment, unless a normalized metric of confidence is used, because the neighborhood 512 of [1e]] the test document 511 is less dense.